

## THE GEOMETRY OF MEANING

Radek Ocelák

GÄRDENFORS, Peter (2014): *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. Cambridge, MA: MIT Press, 360 pp.

Peter Gärdenfors' latest book is a highly important contribution to the theory of meaning in natural language — not to one of the many particular branches of semantic theorizing, but to semantics in general. I take it for given that nowadays, true and interesting progress in semantics is not to be found in works that remain within the narrow theoretical limits of cognitive semantics, formal semantics, psycholinguistics, neurolinguistics, computational, philosophical, or cross-linguistic semantics. Instead, we should seek it in the works combining the key insights from many of these fields, and more. That is definitely achieved in Gärdenfors' new book. Even if the perspective of cognitive linguistics is particularly strong in most of the book, it is substantially enhanced with that of general cognitive science (including computational experiments), computational linguistics, with psychological and psycholinguistic evidence. The author's very name, moreover, guarantees that all sorts of formal, logical and mathematical aspects are taken into consideration as well. His characterization of himself as a giraffe, attempting to comprise the entire semantic savanna in one view (p. xi), is thus more than justified; the amount of relevant literature taken into account is impressive. It should be also mentioned that the book is very thoroughly edited. In my opinion, it constitutes a potential milestone in cognitive linguistics: devising a socio-cognitive, rather than individually cognitive, perspective of meaning, it virtually makes it possible for cognitive linguists and many other linguists as well as philosophers to talk to each other again.

A crucial idea of the book is that there is more structure to natural language meaning than has been traditionally acknowledged in the formal, logical or set-theoretic accounts. This structure is geometric in nature. Objects as well as other components of our experience can be fundamentally characterized in terms of their similarity in various respects or dimensions. Some of the dimensions come hand in hand as domains, which can be represented as geometric spaces of various dimensionality. Concepts that find expression in natural languages then are not "free" to pick just anything from such spaces: they characteristically carve out *regions* that are *continuous*, and moreover, *convex*. That, as a thesis, gives important (and apparently very plausible) predictions as to how the semantics of any human language can and cannot be structured, predictions which cannot be easily obtained in the more traditional formal framework. It also suggests a very efficient way of learning categories from examples, and provides a convincing solution to the question of how it is possible to achieve effective communication at a reasonable cognitive cost.

In his book, Gärdenfors generalizes the basic idea of meanings as regions in conceptual spaces so as to treat a number of familiar cognitive linguistic topics, such as object categories, nouns, properties, adjectives, actions, events, verbs, prepositions,

and metaphors, and less familiar ones, such as forces or compositionality. Some more valuable predictions of the book concern the correlated acquisition of words from a particular domain, or the grouping of metaphors which draw on the structural similarity of particular domains.

All this being said, I would like to argue that in this book, Gärdenfors did not quite manage to keep his position consistent; he does not fully carry out what the first part of the book commits him to. In the second part, he largely recedes from the socio-cognitive position elaborated previously. For me, this furthermore indicates that even the departure from the cognitive tradition undertaken in the first part may not be radical enough. These will be the main critical points of the present review.

The above mentioned laudable step in the socio-cognitive direction consists in the following. We should not *assume* (as the cognitive linguistic tradition generally does) that the mental entities that we regard as the meanings of various expressions are, in some mysterious way, the same for all speakers of the language in question. If the inter-individual sharing of mental imagery (image schemas etc.) is to make any sense at all, it should be explained how this mental convergence can be brought about. Gärdenfors attempts such an explanation in chapter 5. Here, meanings are explained in terms of fixpoints, or equilibria, of communicative interactions. These game-theoretic notions describe states of communication in which the communicators have no incentive to unilaterally change their coding or interpreting strategies, as any such change would make communication less efficient. Assuming that what is communicated are positions in a conceptual space (e.g., we describe the color of an object), this makes clear how the same concept (say, *blue*), as a mental object, can come to be shared inter-individually. Roughly speaking, we simply keep changing our coding and interpreting strategies, until they optimally fit the other's strategies by being basically the same.

Equally importantly, however, the view leaves room for our individual concepts *not* being the same, even if this is less emphasized by the author. A communication fixpoint can be reached even when our coding and interpreting strategies differ from the other's strategies. (It is also worth noting that not all of our communication dwells in a fixpoint stage, from a synchronic point of view. Yet this does not imply that there is no semantics in such a case.) The author is also aware of the option that our concepts may not even "live" in the same space (as is arguably the case, e.g., with the color concepts of the color-blind speakers). In the context of cognitive linguistics, showing how concepts (such as the concept of blue) *can* be inter-individually shared, and that they *need not* be, is in my opinion a major achievement of Gärdenfors and his references.

Note that the employed notion of meanings and concepts is still a fairly mentalistic one; it is not a use-theoretic conception in Wittgensteinian lines. The use of words in communication serves as a means for conceptual coordination, but what is coordinated are unequivocally concepts as mental objects, regions in the conceptual spaces that characterize the mind of each particular individual.

Gärdenfors seems to assume that it is, after all, normal for these concepts to be fully coordinated between individuals, lack of coordination being something extraordinary or defective (an example given (on p. 99) concerns communication

between adults and children). That must be why in most of the book the author feels no need to keep apart the mental level of concepts and the intersubjective level of meanings as communication fixpoints. If our concepts are fully shared, analysis on the intersubjective level is of course omissible for all practical purposes. And indeed, despite the revolutionary chapter 5, most of Gärdenfors' analyses are completely in line with the cognitive linguistic tradition in that no important distinction between concepts (conceived in a mentalistic way) and meanings is exercised. Even where the two levels are distinguished, Gärdenfors finds it convenient to label mentalistic concepts "individual meanings" (p. 18; emphasis RO). The insight that what is individual is *not* meaning yet is apparently not very firmly entrenched. In the second part of the book, the option that our individual concepts may not be fully coordinated is as if forgotten. Semantics of nouns, adjectives or verbs is discussed, building on regions in conceptual spaces, and the implication seems to be that the author claims cognitive adequacy on the level of any single individual. That is surely disappointing for anyone who has been often surprised, as I have, that cognitive linguists can know how *love* or *climbing* is conceptualized in our minds without having actually examined the mind of each and every one of us.

In the book, we find several harmless references or allusions to the later work of Ludwig Wittgenstein. Gärdenfors even characterizes his semantic theory as a combination of cognitive linguistics, conceptual spaces, and *language games* (p. 265). The truth however is that Wittgenstein would emphatically reject the mentalistic view of meanings or concepts as necessarily involving definite objects and recurring processes in the mental realm, mental objects and processes which would regularly correspond to the expressions of language. He would favor neither Gärdenfors' placing of various conceptual spaces into the minds of individuals, nor the idea that having concepts is a matter of carving regions out of such spaces mentally.

In my own view, locating conceptual spaces in individual minds and conceiving them as shared backgrounds for establishing shared concepts can be rather safe, but only for the most basic, biologically well-grounded domains. For instance, it is quite plausible to assume that the visual perception by each of us (except the fraction of color-deficient observers in the population) induces the same similarity-based color space (although its characterization by the "color spindle" (p. 23) is questionable). It then makes enough sense to describe color concepts as regions within that space, and to regard such descriptions as statements concerning our individual cognition. Here, the mentalism characteristic of the cognitive linguistic tradition seems by and large adequate, or at least relatively easily reconcilable with more use-theoretic perspectives.

I however believe that Gärdenfors is wrong when he decides to make this into a general model of semantics. That, in my opinion, rather quickly turns into a sort of mental metaphysics, into postulating mental structures for which there is no evidence and for which it is hard to imagine any. The problem is thus not merely that little research has been done so far, as the author suggests in some places. With the exception of some very basic domains (such as size, color, or temperature), the semantic description in geometric terms requires many-dimensional spaces, spaces where the meaning of particular dimensions is hard to determine, and higher-order spaces which provide, e.g., for the geometric construction of events out of

simpler components. For the more complex of these structures, there seems to be no other motivation than to provide something in the mind of every single individual which would correspond to various language expressions. For example, the mental representation suggested for the event of Oscar pulling a sledge to the top of the hill (pp. 160–161), or for the sentence describing this event, is highly complex. It is doubtful that more individuals could arrive at an identical one, or what it would even mean for them to have done so. The author emphasizes in several places that what he aims for is a cognitive, and not a scientific interpretation of dimensions, forces, events etc. Yet it seems very clear that some of his analyses are solely motivated by what the mathematics of his conceptual tools offers. For instance, the solution of telicity vs. atelicity (p. 172ff.), the discussion of compositionality (p. 241ff.), or the view that states are limit cases of events (p. 163), and objects limit cases of concepts (p. 127), seem very *ad hoc* in this sense.

I decidedly believe that explaining meanings across the board as regions in conceptual spaces is a fruitful way of looking at natural language semantics. What I do *not* believe is that it can plausibly get a straightforward mentalistic interpretation, with the exception of the most basic cognitive domains. Rather, I take it as a claim concerning the structures of the intersubjective rationality, with which we individually struggle, using our cognitive resources. I think most of the conceptual spaces relevant for our language and thinking are intersubjective constructions: they are constituted, *inter alia*, by the appropriate use of certain expressions, which many of the individual speakers may never fully master on their own. (Take for instance the space of political positions: left, right, fascist, liberal, libertarian, conservative...) It therefore does not make much sense to conceive the meanings in these spaces as supervenient on individualistic, mental ones. In contrast, we may say that Gärdenfors, like many other authors in the cognitive tradition, subscribes to an unambiguous solution to the chicken-or-the-egg problem of individual cognition and language meaning: the former is the chicken and comes first, and this is supposed to hold across the board.

True, the author's analysis is largely limited to the more concrete, cognitively more basic domains, which "are required for the development of communication during a child's first years" (p. 24) and for which the mentalistic interpretation may seem relatively adequate. Even if this were the intended significance of the book, however, the problem would be that Gärdenfors does not discuss where the mentalistic view stops being tenable, or how we can add abstract domains to the overall picture, without creating a dubious gap between the more concrete and the more abstract domains of semantics.

Neither is it convincingly shown that the boundary is not seriously trespassed in some of the analyses. For instance, we might subscribe to the view that Oscar's pulling of the sledge can be in some sense very rationally represented by a force vector and a result vector in some abstract underlying domain, while denying that this is what each of us performs in his or her mind on each and every occasion when Oscar's action is discussed. In a similar fashion, we can question the mechanism of mental focusing, which plays an important role for Gärdenfors as well as for other works in cognitive linguistics. It is surely a true point that a sentence describes an

event focusing on some of its aspects, or that a sentence in the passive voice focuses on the theme rather than on the agent. Only, we should not claim that such focusing is necessarily something that people do in their minds when using these sentences. For how would we know?

I cannot refrain from pointing to a particular place of Part II, where the perspective of intersubjective rationality forces itself into the book as if against the author's will. It is claimed (p. 126) that the sentence *An abyssinian is a cat* "will automatically be true by the fact that the regions associated with the domains for the category of an abyssinian are subregions of those associated with the category of a cat". Such references to the notion of *truth* are rare in cognitive linguistics, and it is for a good reason. Namely, as far as the individual concepts of any of us are concerned, abyssinians may as well fall in the category of horses or mushrooms. The fact that they *are* cats (and not horses or mushrooms), or that the quoted sentence is *true*, importantly transcends the level of individual cognition.

Gärdenfors' book, serious, thorough and important as it is, is in my opinion problematic in that it does not define what might prove its cognitive proposals wrong. A particular problem that I see consists in the suggested principles of contextual dependence, of varying weights of the domains involved in particular concepts, as well as in the changes in mental focusing. My worry is that these theoretical devices might actually serve to secure the theory from possible falsification. It would be highly unfortunate if conceptual spaces were to turn into an omnivorous and irrefutable way of thinking, rather than an actual scientific theory.

**Radek Ocelák** | Institute of Philosophy and Religious Studies, Faculty of Arts, Charles University in Prague  
<radioc@seznam.cz>